



An artificial intelligence decision support system for the management of type 1 diabetes

Nichole S. Tyler¹ , Clara M. Mosquera-Lopez¹, Leah M. Wilson² , Robert H. Dodier¹, Deborah L. Branigan², Virginia B. Gabo², Florian H. Guillot², Wade W. Hilt¹, Joseph El Youssef², Jessica R. Castle² and Peter G. Jacobs¹  

Type 1 diabetes (T1D) is characterized by pancreatic beta cell dysfunction and insulin depletion. Over 40% of people with T1D manage their glucose through multiple injections of long-acting basal and short-acting bolus insulin, so-called multiple daily injections (MDI)^{1,2}. Errors in dosing can lead to life-threatening hypoglycaemia events (<70 mg dl⁻¹) and hyperglycaemia (>180 mg dl⁻¹), increasing the risk of retinopathy, neuropathy, and nephropathy. Machine learning (artificial intelligence) approaches are being harnessed to incorporate decision support into many medical specialties. Here, we report an algorithm that provides weekly insulin dosage recommendations to adults with T1D using MDI therapy. We employ a unique virtual platform³ to generate over 50,000 glucose observations to train a *k*-nearest neighbours⁴ decision support system (KNN-DSS) to identify causes of hyperglycaemia or hypoglycaemia and determine necessary insulin adjustments from a set of 12 potential recommendations. The KNN-DSS algorithm achieves an overall agreement with board-certified endocrinologists of 67.9% when validated on real-world human data, and delivers safe recommendations, per endocrinologist review. A comparison of inter-physician-recommended adjustments to insulin pump therapy indicates full agreement of 41.2% among endocrinologists, which is consistent with previous measures of inter-physician agreement (41–45%)⁵. In silico^{3,6} benchmarking using a platform accepted by the United States Food and Drug Administration for evaluation of artificial pancreas technologies indicates substantial improvement in glycaemic outcomes after 12 weeks of KNN-DSS use. Our data indicate that the KNN-DSS allows for early identification of dangerous insulin regimens and may be used to improve glycaemic outcomes and prevent life-threatening complications in people with T1D.

Optimal management of T1D requires precise insulin administration to maintain glucose within safe ranges. Dosage regimens are complicated by day-to-day changes in insulin sensitivity, which cause large excursions in glucose. Failure to dose insulin properly can result in diabetic ketoacidosis and hypoglycaemia, which may lead to coma or death. Intensive insulin regimens can enhance glycaemic outcomes in people with T1D who use MDI therapy⁷, but a number of factors confound adherence to insulin dosing. Fear of hypoglycaemia, challenges with numeracy to calculate meal or correction boluses, changes in insulin sensitivity during exercise, illness, stress and menstruation, and the psychological toll of this chronic disease make it difficult for people with T1D to adhere to these regimens^{8–11}.

Whereas many smartphone apps are available to help people better manage their diabetes, most of these are not validated and have not shown clinical efficacy. A recent review indicated that out of hundreds of such applications, only 12 were validated in clinical trials and few of these significantly improved glycated haemoglobin (HbA1c) in people with T1D^{12–15}. Apps shown to improve glycaemic outcomes provided users with weekly or biweekly feedback from health professionals on insulin dosage adjustments^{12,13}. Continuous glucose monitoring (CGM) has been shown to significantly improve HbA1c, but as a sole intervention does not bring everyone to goal¹⁶. CGM-informed advisory systems¹⁷ range from machine learning to physiological models and heuristic approaches for the adjustment of basal^{18–20} and bolus therapies^{21,22}. Nimri et al.⁵ developed a system to guide adjustment of insulin pump settings using capillary blood glucose or CGM data. Perez-Gandia et al.²³ developed a predictive neural network to assist with real-time insulin administration or carbohydrate consumption. MDI-inclusive approaches include a model-based decision support system for titration of insulin prior to exercise by Breton et al.²⁴, and an adaptive KNN case-based reasoning approach for titration of short-acting insulin by Reddy et al.²⁵.

The KNN-DSS that we describe provides up to four optimally selected dosing and behavioural recommendations once per week to adults with T1D who use MDI therapy. Recommendations are selected to manage insulin dosed for meals and snacks to bring glucose to within a target range. A virtual patient simulator platform was implemented to design the KNN-DSS algorithm. The virtual patient simulator³ is a mathematical representation of the glucoregulatory response to food, insulin and exercise in people with T1D. The virtual patient simulator was used to train a machine-learning KNN⁴ model to predict optimal insulin recommendations that improve glycaemic outcomes. Input data for the algorithm are acquired from CGM data, insulin data obtained from Bluetooth-enabled capture devices and physical activity metrics obtained through wearable sensors (Fig. 1a). The KNN-DSS then classifies glycaemic features and delivers recommendations to improve percent time in target range (70–180 mg dl⁻¹) and reduce percent time in hypoglycaemia (Fig. 1b). User-specific titration of insulin occurs using an adaptive learning postprandial hypoglycaemia avoidance (ALPHA) algorithm²⁶ which selects the optimal bolus insulin based on the prior glycaemic outcomes of the user (Fig. 1c). To ensure that recommendations conform to physician standards, we developed an expert-knowledge quality control algorithm (Fig. 1d, Extended Data Figs. 1–7). The heuristic quality control algorithm is designed for user safety and operates independent of the machine-learning framework. The KNN-DSS system delivers

¹Artificial Intelligence for Medical Systems (AIMS) Lab, Oregon Health & Science University, Portland, OR, USA. ²Harold Schnitzer Diabetes Health Center, Oregon Health & Science University, Portland, OR, USA. ✉e-mail: tylern@ohsu.edu; jacobsp@ohsu.edu

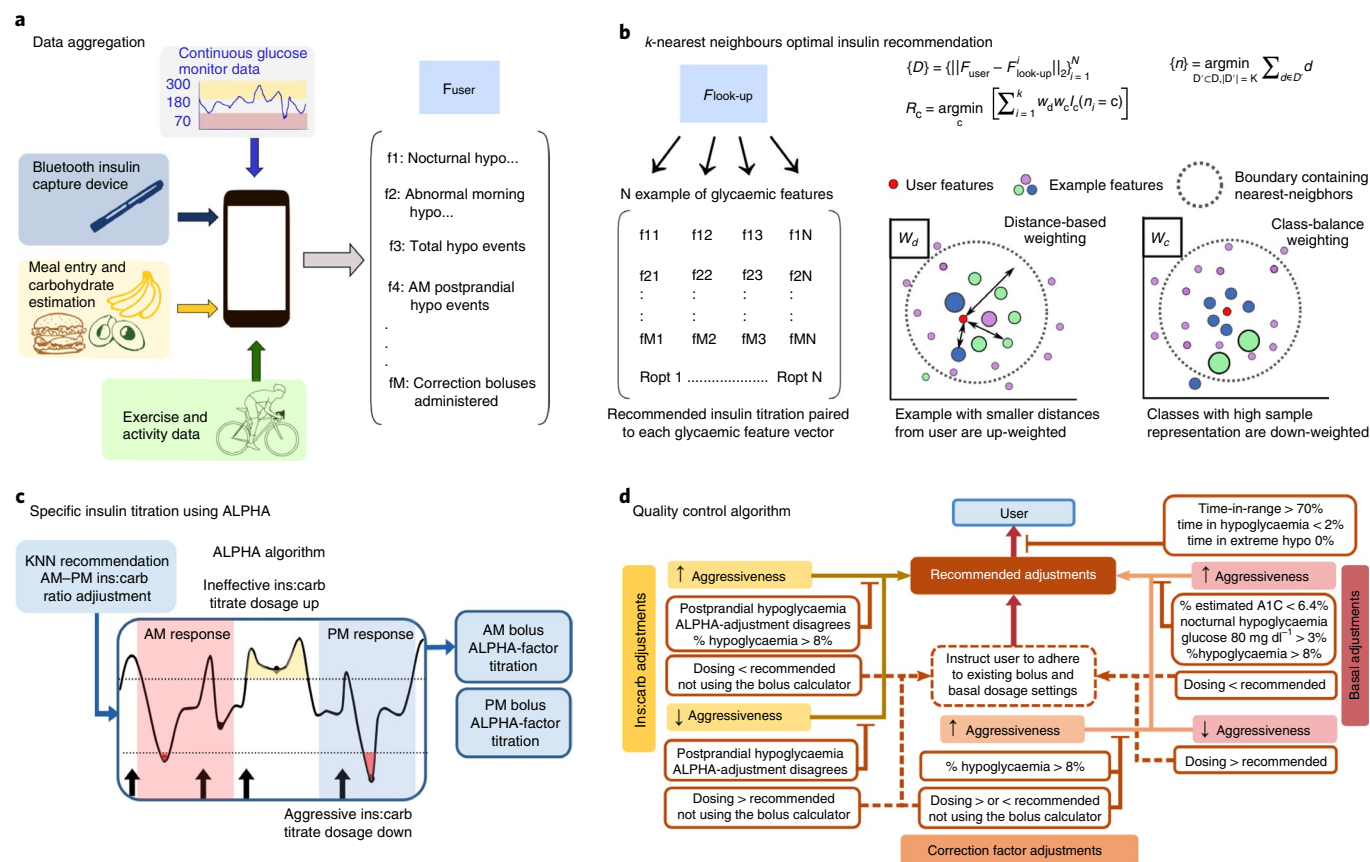


Fig. 1 | Decision support engine framework to identify user-specific insulin titrations. **a**, The user data are aggregated and processed for extracting glucose, insulin, meal and exercise features that may be used to optimally titrate insulin doses. **b**, The user features (F_{user}) are matched to the closest examples in the look-up table for the k -nearest neighbours algorithm, $F_{\text{look-up}}$. The distance between user features and the examples in the look-up table are calculated as $\{D\}$. The k examples within minimum distance to user features, $\{n\}$, are weighted by distance, w_d , and class-size, w_c ; the final insulin dosage recommendations, R_c , are returned by the KNN algorithm. **c**, For those recommendations indicated by the KNN-DSS, the ALPHA algorithm assigns an aggressiveness factor that titrates carbohydrate ratios and correction factors to improve time in target range and reduce time in hypoglycaemia. **d**, A quality control algorithm is employed to ensure that KNN-DSS recommendations adhere to physician standards.

one or more recommendations from a set of 12 unique recommendations for insulin adjustments and dosage behaviours with respect to long-acting basal insulin, fast-acting carbohydrate-to-insulin (carb:ins) ratio, and correction insulin dosage (Table 1). The top three meal and basal insulin recommendations are selected by KNN classification, whereas recommendations for correction doses and compliance with care are supplied by the heuristic ALPHA and quality control algorithms.

We validated the accuracy and safety of KNN-DSS-generated recommendations compared with those of board-certified endocrinologists using 687 days of real-world data collected from 25 adult participants on MDI therapy. We demonstrated efficacy of the KNN-DSS during two 52-week studies in silico and characterized the response of the engine to dynamic disturbances in glycaemic patterns. Lastly, we report the results of a short, proof-of-concept, single-centre clinical study to evaluate the safety of the KNN-DSS in human participants (Supplementary Table 1: Study 1).

To compare the recommendations of the KNN-DSS with endocrinologist recommendations, we used data collected from 25 adult participants with T1D during a 28-d outpatient study (Supplementary Table 1: Study 1). One of three endocrinologists from the Oregon Health and Science University (OHSU) medical centre analysed the glucose and insulin dosing data from each participant and provided recommended adjustments to basal insulin and to fast-acting meal insulin and correction insulin during four

different windows of time (7:00–11:00, 11:00–15:00, 15:00–20:00 or 20:00–7:00) (Table 1). Recommendations regarding the daily bolus calculator use were also provided. The KNN-DSS recommendations were labelled as being in full agreement, partial agreement, full disagreement or partial disagreement with the physician. The accuracy of recommendations delivered by the KNN-DSS as compared to those of board-certified endocrinologists was quantified using a modified Sørensen–Dice coefficient (DSC)²⁷ (equation (6)). We measured a combined agreement of 67.9% between endocrinologist and KNN-DSS recommendations, whereas 6.4% of recommendations were in disagreement. In 16.7% of recommendations the engine identified an issue and the physician did not indicate a recommendation, and 9% of recommendations were not comparable (Table 2). We performed additional analysis on a subset of these participants who exhibited consistent use of the bolus calculator and adherence to their insulin dosage settings (greater than 75%). For this subset of participants, the engine recommendations were in full agreement with those of physicians 50.8% of the time and exhibited an overall agreement of 67.5% with the recommendations of physicians. We observed that over 99% of recommendations delivered across 100 weeks of data by the KNN-DSS passed a safety review in which the endocrinologists reviewed each recommendation for the potential to cause hypoglycaemia episodes and overnight events.

The measure of physician agreement with the KNN-DSS is similar to those found in published studies involving other decision

Table 1 | Recommendations delivered by the KNN-DSS engine

Recommendation	Message to user	Adjustment window	Titration method
Basal adjustment	You may need (less/more) basal insulin. It is recommended that you (decrease/increase) your AAA insulin from BBB to CCC units.	AM PM	Adjustment by $\pm 10\%$ from weekly baseline settings
Carb:ins ratio	You may need (less/more) insulin before (breakfast/lunch/dinner/an evening meal). It is recommended that you change your carb ratio from AAA to BBB.	7:00–11:00 11:00–15:00 15:00–20:00 20:00–7:00	ALPHA algorithm: meal bolus glycaemic response and assignment of dosage titration.
Correction factor	You may need (less/more) insulin in (morning/afternoon/evening/night-time) to bring down high glucose levels. It is recommended that you change your correction factor from AAA to BBB.	7:00–11:00 11:00–15:00 15:00–20:00 20:00–7:00	ALPHA algorithm: correction bolus glycaemic response and assignment of dosage titration.
Bolus adherence ^a	You may have taken (more/less) insulin than recommended by the bolus calculator during certain times of day. It is recommended that you use the amount recommended by the bolus calculator.	All day	N/A
Basal adherence ^a	We have found that the amount of basal insulin that you are taking is different than the amount we recommend. Taking the recommended amount may improve your glucose levels.	All day	N/A

Recommendations for insulin dosage are titrated to be higher or lower during different time windows using the specified titration method. ^aBehavioural recommendation.

support systems. In an international, multicenter study, Nimri et al.⁵ compared physician-recommended adjustments of insulin pump settings and found the expected full agreement of different physicians to be between 41% and 45%, and the expected full disagreement between 9% and 12%. The results demonstrated high variability among physicians both internationally and in practice at the same institution. Nimri et al.⁵ used multiple study centres and evaluated insulin pump settings, which differentiated their study from that of a single-centre study on people using MDI. Additional studies that compared adjustments in MDI therapy reported that general practitioners and endocrinologists alike identified 67% of indicated changes to insulin²⁸, and reported an agreement of 63% between physician-recommended and software-recommended titrations to insulin dosages²⁹.

We measured inter-physician recommendation variability on a dataset collected previously from participants with T1D using sensor augmented pump therapy³⁰ (Supplementary Table 1: Study 2). We found a value of 41.2% in relation to full agreement among endocrinology faculty at OHSU (Supplementary Table 2). The KNN-DSS engine, trained with a virtual platform, demonstrates high full agreement (50.8%) and partial agreement (67.5%) with endocrinologist recommendations when validated on real-world data (Study 1), and exceeds the inter-physician agreement found in Study 2.

We evaluated the ability of the KNN-DSS to improve glycaemic outcomes using two virtual patient simulators^{3,6}. Each virtual patient was given real-world meal scenarios previously recorded during a clinical trial of automated insulin delivery therapy. These meal scenarios are meant to rigorously challenge algorithm performance with realistic eating patterns.

Evaluation in silico demonstrated the ability of the KNN-DSS to identify problematic glycaemic patterns and to deliver effective insulin dosage recommendations. In the first in silico study using the OHSU T1D simulator³, 29 virtual patients with varying adherence to insulin therapy, varying circadian insulin sensitivities and carbohydrate misestimation of $\pm 30\%$, were evaluated in a

75-week study of weekly decision support. After 12 weeks of use, the KNN-DSS and supporting algorithms improved virtual patient outcomes considerably, increasing average percent time in target range from 59.5% to 79.8% ($P = 2 \times 10^{-5}$), maintaining percent time in hypoglycaemia at target ($< 2\%$) and reducing inter-individual variability (Fig. 2a and Supplementary Table 3).

After the optimal insulin dosing settings were obtained at 52 weeks of simulation, we disturbed the system by imposing new insulin dosing errors and changes to patient settings (Supplementary Table 4). The engine was able to correct the problems with the invalid dosing settings, gradually improve the time in target range, and reduce the time in hypoglycaemia (Fig. 2b). Patient settings largely trended towards their pre-disturbance values (Fig. 2c), which indicates the engine's ability to respond to dynamic changes and retrieve the original settings. Since there are many possible combinations of long-acting and short-acting insulin therapy that may improve glycaemic control in a person with T1D, not all therapy settings returned to the original settings prior to the disturbance. We nonetheless found that all of the glycaemic outcomes of the virtual patients still improved following the disturbance (Fig. 2b).

In the second 52-week in silico study, we used a benchmarking platform accepted by the United States Food and Drug Administration for the evaluation of artificial pancreas algorithms, the UVA-Padova simulator⁶, to evaluate the performance of the KNN-DSS on 100 virtual adult patients, 100 virtual adolescent patients and 100 virtual paediatric patients. In adult patients, percent time in target range improved from 75.1% at baseline to 81.8% ($P = 1 \times 10^{-4}$) at the study conclusion. Percent time in hypoglycaemia was reduced from 4.0% at baseline to 0.55% ($P = 2 \times 10^{-12}$) at the end of the study (Supplementary Table 3 and Extended Data Fig. 8). In silico studies may provide optimistic estimations regarding glycaemic outcomes because virtual study patients exhibit near perfect adherence to recommended insulin adjustments, which often does not happen in real-world studies. Nonetheless, the results shown here indicate that study participants who use the recommendations provided by the KNN-DSS may expect a reduction

Table 2 | Agreement between KNN-DSS and endocrinologist recommendations using real-world human data

Recommendation comparison	% Agreement	% Disagreement	% Additional	% Not comparable
Assessing recommendations on all participant data. (<i>n</i> = 78 weeks)	Full 27.8 Partial 40.1 Overall 67.9	Full 0.4 Partial 6.0 Overall 6.4	Physician 6.4 Engine 10.3 Overall 16.7	9.0
Assessing recommendations on participant data with > 75% recommendation adherence (<i>n</i> = 39 weeks)	Full 50.8 Partial 16.7 Overall 67.5	Full 2.6 Partial 8.8 Overall 11.4	Physician 13.2 Engine 7.9 Overall 21.1	0.0

Agreement is calculated using a Sørensen–Dice coefficient similarity comparing physician recommendations to engine recommendations for each week of observed data. Adherence refers to participant use of the bolus calculator.

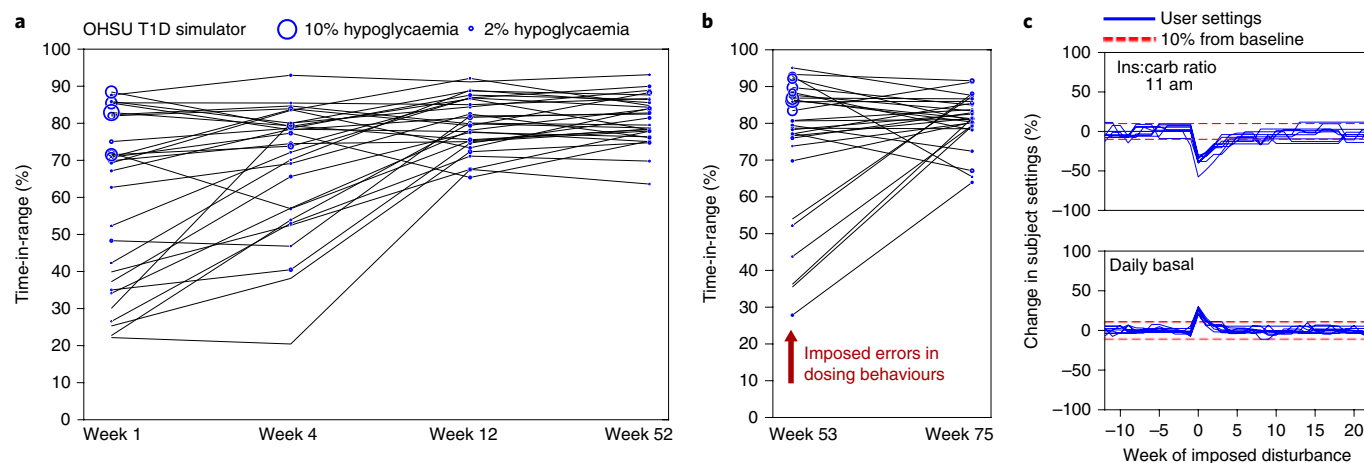


Fig. 2 | Engine performance in improving patient outcomes in silico. a, Outcomes of the in silico evaluation of the KNN-DSS over 52 weeks. Virtual patients from the OHSU T1D simulator undergo weekly use of the KNN-DSS engine. **b,** At 52 weeks, new insulin settings and dosing behaviours are imposed on patients, the effects of which are measured at 53 weeks. **c,** Evolution of patient settings 20 weeks from baseline following a disturbance at week 52. The dashed red line indicates 10% from baseline.

in hypoglycaemia and an improvement in time in target range after 12 weeks. Notably, at 4 weeks, both the UVA–Padova and the OHSU T1D simulators showed reductions in glycaemic variability and hypoglycaemia and a small average improvement in glycaemic time in target range, with many patients showing a reduction in time in target range. The KNN-DSS is designed to prioritize safety and reduce hypoglycaemia as well as optimize insulin dosages and percent time in target range; as a consequence, time in target range may initially be reduced as problematic and aggressive insulin doses are titrated down, and improves significantly after 12 weeks of continued use. Performance was also evaluated using the UVA–Padova simulator in an adolescent and paediatric population. For adolescents and children, we also observed significant improvements in glycaemic outcomes, as percent time in target range increased from 68.2% to 75.2% ($P = 8 \times 10^{-9}$) for adolescents, and 65.7% to 69.1% ($P = 0.02$) for children after 12 weeks (Supplementary Table 5).

We evaluated the safety of the KNN-DSS in a single-centre feasibility study. Sixteen adults out of 25 adults with type 1 diabetes on MDI therapy (Supplementary Table 1: Study 1) underwent weekly KNN-DSS augmented decision support. These 16 participants were given weekly recommendations on dosing and behavioural changes from a physician who reviewed the glucose history of participants and recommendations suggested by the KNN-DSS. Although this small pilot study was not powered to detect any impact from the intervention, and the study duration of 4 weeks was too short to observe a significant impact on time in target range, we did observe that participants exhibited a 25% decrease in hypoglycaemia events over a 24-h period (from 0.86 to 0.64 events per day, $P = 0.051$), a 33% decrease in daytime hypoglycaemia events (from 0.43 to 0.29 events per

day, $P = 0.096$), a 43% decrease in hypoglycaemia overnight (from 0.50 to 0.29 events per day, $P = 0.04$) and an 76% decrease in serious hypoglycaemia ($<54 \text{ mg dl}^{-1}$) overnight (from 0.48% to 0.11%, $P = 0.03$) during the final week compared with the first week in the study (Extended Data Fig. 9). Fifteen of the 16 participants completed all 4 weeks of the study, whereas one participant chose to exit the study after 2 weeks. In a similar manner, we observed no substantial changes in daytime percent time in target range comparing the final week to week 1 (54.2% to 52.6%, $P = 0.63$), a decrease in time in target range overnight (56.7% to 45.7%, $P = 0.034$), a non-significant decrease in time in target range over a 24-h period (55.3% to 49.1%, $P = 0.06$), and a non-significant increase in mean glucose (172.4 mg dl^{-1} to 185.5 mg dl^{-1} , $P = 0.08$). As shown in Supplementary Table 3, our simulation studies confirm the findings of the clinical study, which indicates that, although a reduction in hypoglycaemia can be expected after four weeks, a longer study is required before we can expect substantial improvements in time in target range.

Other groups have also reported on the impact of automated decision support systems on glycaemic outcomes, with results primarily indicating that hypoglycaemia can be reduced. Breton et al.²⁴ recruited 24 participants with T1D (8 of whom were on MDI therapy) for a crossover study to evaluate automated decision support. Glycaemic outcomes were evaluated during a 48-h in-patient session in which participants underwent standardized meals and exercise. They observed a statistically significant reduction of percent time in hypoglycaemia for the DSS versus the control group (3.2% (1.3, 4.8) versus 0.9% (0.4, 2.3), $P = 0.02$) but no significant change in percent time in target range. Herrero et al.²² described a case-based reasoning DSS (named ABC4D) and evaluated the system in silico

using 20 representative adults and adolescents from the UVA–Padova simulator population. They found that after 4 weeks, percent time in target range could be improved in adults from 75.2 ± 11.7 to 81.9 ± 13.4 ($P < 0.05$) and percent time in hypoglycaemia could also be reduced from 0.3 ± 0.5 to 0 ($P = 0.17$). Although we also used the UVA–Padova population, a direct comparison of ABC4D to the KNN-DSS would be difficult as we evaluated our algorithm in 300 UVA patients. Reddy et al.²⁵ further evaluated the ABC4D algorithm in a real-world, 6-week pilot study, but whereas both percent time in target range and percent time in hypoglycaemia improved slightly, the changes were not significant. Other approaches described a reduction in hypoglycaemia after 12 weeks of basal and fast-acting insulin decision support, but no significance measures were reported⁵. A common theme of these studies is that it has not yet been shown that a DSS can improve percent time in target range in human studies. Our in silico results indicate that longer study durations (>12 weeks) will be necessary to demonstrate improvements in percent time in target range through use of the KNN-DSS.

The KNN-DSS was trained and validated for use with specific sensor technologies, insulin therapies and target populations. The KNN-DSS system uses CGM devices that sample glucose at 5-min intervals. Flash glucose systems are compatible with the KNN-DSS algorithm; however, these will require further testing to handle the asynchronous sample rates of flash glucose systems. Like flash glucose monitoring systems, fingerstick glucose is also measured at varying intervals (4–10 times per day), and more training and testing is needed before this could be incorporated into the KNN-DSS. The engine is also compatible with the large majority (~95%) of existing insulin therapies including fast-acting aspart or lispro and long-acting Lantus (glargine), Tresiba (degludec), and Toujeo (glargine U300) basal formulations. Intermediate-acting NPH insulin (which represents <5% of use cases) will require additional testing and evaluation. Although we report on adults, adolescents and children in this article, further work will be needed to assess performance in vulnerable and complex populations, including the elderly and pregnant women. Moreover, virtual simulators have not yet been developed to fully represent these populations, which makes it challenging to incorporate these populations into the design. The KNN-DSS will need further training before it can be targeted to these groups.

We explored how specific glucose and insulin features were related to optimal recommendations calculated by the KNN-DSS algorithm (Supplementary Table 6). We observed that the KNN-DSS mapping of specific glycaemic features to optimal recommendations matches intuition and, in general, is synonymous with physician opinions regarding titration of insulin dosing for people with T1D.

We have shown that an artificial intelligence decision support system with an expert-knowledge quality control algorithm can be used to help people with T1D identify problematic glycaemic patterns at the same level of accuracy as that of board-certified endocrinologists. Our unique in silico training platform enables us to generate training sets of diverse glycaemic profiles from the OHSU T1D simulator. The final engine design performs well on independent in silico virtual populations and, most notably, can identify insulin dosage issues in real-world human data. Further validation in longer clinical trials in humans is critical to understand how artificial intelligence-based decision support systems can improve glycaemic outcomes in people with T1D.

Methods

k-nearest neighbours design. The *k*-nearest neighbour classification algorithm (KNN)⁴ is a supervised machine-learning approach that matches input features with an outcome variable or class. We define the KNN input features as specific glycaemic outcomes, such as percent time in target range (70–180 mg dl⁻¹), percent time spent in hypoglycaemia (<70 mg dl⁻¹), the number of meal- or correction-related hypoglycaemia episodes and so on (see Supplementary Table 7 for complete list of features). The outcome variable or class that the KNN predicts

is a recommended adjustment to insulin dosage that leads to an improved percent time in target range and a reduction in percent time in hypoglycaemia. This classification is accomplished using a look-up table that matches the weekly glycaemic features of a person with their optimal dosing recommendations. The KNN approach identifies unique recommendations regarding long-acting basal insulin and fast-acting meal insulin. Recommendations regarding adherence and correction factors are accomplished separately by the ALPHA and quality control algorithms described in detail below (Fig. 1).

Training dataset generation. We generated the look-up table of optimal recommendations using an in silico virtual patient simulator consisting of 99 individuals with T1D exhibiting diverse glycaemic dynamics generated by variations in insulin sensitivity and daily insulin requirements, carbohydrate sensitivity, weight differences, circadian insulin sensitivity and variations in adherence to insulin dosing³. Circadian insulin sensitivity was achieved through continuous modulation of insulin-mediated peripheral glucose uptake, peripheral insulin uptake and hepatic glucose production, varying within 20% of their original values³. CGM data and insulin data were obtained from 70 of the virtual patients over the course of a 15-week in silico study, whereas the other 29 patients were retained for an evaluation of algorithm performance in a separate in silico study. Real-world meal scenarios provided to the virtual patients were obtained from a previous clinical trial¹¹ in which we acquired 80 d of realistic meal patterns and carbohydrate content. Forty of these daily meal scenarios were used to train the KNN-DSS, while the other 40 were used to validate the algorithm. These daily meal scenarios were randomized and administered to virtual patients, and the Pettus–Edelman³² approach for CGM trend arrow adjustment and a bolus calculator were used to dose mealtime insulin. The optimal recommendations for an individual were identified by simulating the glycaemic outcomes from administering each dosing recommendation given in Table 1. The glycaemic outcomes of time in glucose target range (70–180 mg dl⁻¹), mean glucose, hypoglycaemia (<70 mg dl⁻¹), and serious hypoglycaemia (<54 mg dl⁻¹) were measured the week following each simulated dosing scenario. During the training of the KNN, these measured outcomes were used to define a heuristic objective function to select the optimal recommendations, $\{R_{opt}\}$, to maximize the measured percent time in target range, and reduce the percent time in hypoglycaemia (X_{TIR} and X_{hypo} in equation (1), respectively). We provide this heuristic objective function in equation (1). In step 1, we identify a subset of recommendations, $\{R_{hypo}\}$, that yields percent time in hypoglycaemia less than or equal to 2% and no serious hypoglycaemia. If the patients exhibit persistent hypoglycaemia, we identify the recommendations that minimize hypoglycaemia, $\{R_{hypo}\}$. Recommendations that yield a percent time in hypoglycaemia greater than 2%, or any extreme hypoglycaemia, are excluded from the subset of optimal recommendations. In step 2, we identify from this subset the optimal recommendation, $\{R_{opt}\}$, which yields the largest improvement in the time spent in target glucose range, or mean glucose if the percent time in target range is too small due to persistent hyperglycaemia.

$$R_{Opt} = \begin{cases} \text{Step 1a, } \{R_{hypo}\} = \cup_{i=1}^n i * \delta(x_{hypo}) * \gamma(x_{serious\ hypoglycemia}) & \delta(a) = \begin{cases} 0, & a > 2 \\ 1, & a \leq 2 \end{cases} \\ \text{Step 1b, } \{R_{hypo}\} = \operatorname{argmin}_{x_{hypo}} \sum_{i=1}^n x_i & \gamma(a) = \begin{cases} 0, & a \neq 0 \\ 1, & a = 0 \end{cases} \\ \text{Step 2a, } \{R_{opt}\} = \operatorname{argmax}_{x_{TIR}} \sum_{i=R_{hypo}} x_i & \rightarrow \text{perform if step 1a yields empty set} \\ \text{Step 2b, } \{R_{opt}\} = \operatorname{argmin}_{x_{MeanGlucose}} \sum_{i=R_{hypo}} x_i & \rightarrow \text{perform if step 2a yields empty set} \end{cases} \quad (1)$$

This optimal recommendation was stored with the glycaemic features from the prior week to form an observation in the look-up table. The observation of paired weekly glycaemic features and optimal insulin dosage were compiled into a look-up table used in the KNN algorithm. Additional real-world behavioural scenarios were imposed during the in silico study by programming the virtual patient to perform one or more of 13 insulin dosing errors (Supplementary Table 4), to periodically accept or fail to accept the advice of the recommendation given each week and to use of two different types of bolus calculators. The size of the final look-up table totalled 51,831 observations.

k-nearest neighbours parameter identification and feature selection. The features used in the KNN-DSS included features drawn from CGM, physical activity data, and long-acting and short-acting insulin data. We identified an optimal set of features through a feature selection technique called ‘greedy’ sequential forward selection³³. We determined the optimal number of neighbours by performing a grid search using the optimal set of features. The final KNN design used 30 neighbours and 25 glycaemic features to perform classification (see Supplementary Table 7). The weighting scheme was decided by comparing classification accuracy across (1) no weighting, (2) distance-based weighting,

(3) class-based weighting and (4) combined distance and class-based weighting (Supplementary Tables 8 and 9).

Feature importance. We evaluated which features contributed most significantly to each recommendation by calculating the ‘mutual information’ between each feature and the recommendation selected by the classifier. The mutual information between two random variables (for example, a feature f with a distribution F and a recommendation r with a distribution R) with a joint probability mass function (P_{FR}) is defined according to equation (2).

$$I(F, R) = \sum_{f,r} P_{FR}(f, r) \log \left(\frac{P_{FR}(f, r)}{P_F(f)P_R(r)} \right) \quad (2)$$

The mutual information of features was calculated for each separate recommendation in the classifier. A one-versus-all approach was used, in which the mutual information considers a single recommendation class as positive and all other classes as negative. This was repeated for each recommendation class in order to generalize what features contribute most significantly to a given recommendation class. Relative feature importance determined for each recommendation is listed in Supplementary Table 6.

Precise insulin titration. The KNN-DSS assumes that meal insulin doses are calculated using the carb:ins ratio and that correction boluses are calculated using correction factors and glucose trends^{32,34}. The system also uses smart-bolus calculations that incorporate mealtime corrections and active insulin on board (IOB) (equation (3)).

$$\begin{cases} \text{Recommended insulin bolus} = \text{bolus}_{\text{meal}} + \text{bolus}_{\text{correction}} - \text{IOB} \\ \text{Bolus}_{\text{meal}} = \text{grams of carbohydrate} \times \frac{1 \text{ unit of insulin}}{\text{carb:ins}} \\ \text{Bolus}_{\text{correction}} = (\text{current glucose} + \text{offset} - \text{target glucose}) \times \frac{1 \text{ unit of insulin}}{\text{correction factor}} \end{cases} \quad (3)$$

Precise titration of insulin is accomplished using heuristic approaches that adjust carb:ins ratios and correction factors. The ALPHA algorithm, described in our recent publication²⁶, retrospectively analyses the average glycaemic response of a person to insulin boluses and returns an aggressiveness factor (A_f). The framework of this algorithm, as well as the modifications to the original algorithm for the current paper implementation, is described as follows. Separate analysis is performed for both meal-related boluses and correction boluses, again across different windows of time (see Table 1). For each bolus entry, ALPHA adapts the aggressiveness factor if the glucose of the person is outside of the target range following meals or corrections. The aggressiveness factors corresponding to each bolus are then used to calculate an average, smoothed aggressiveness factor (A_f^{avg}) shown in equation (4). This smoothed aggressiveness factor is used to adjust carb:ins ratio and correction factor settings.

$$A_f^{\text{avg}}(k) = \frac{A_f(k) + A_f^{\text{avg}}(k-1) + A_f^{\text{avg}}(k-2)}{3} \quad (4)$$

In the implementation discussed herein, the aggressiveness factor assigned to an individual bolus delivered at time k , $A_f(k)$, is determined using a piece-wise linear adjustment that is a function of the minimum glucose ($G_{\min}$) measured within 4 h of the last meal bolus or within 3 h of the last correction bolus (equation (5)).

$$A_f(k) = \begin{cases} 0.4, & 0 \leq G_{\min} \leq G_{\text{hypo}} \\ \frac{(G_{\min} - G_{\text{hypo}}) \cdot (A_f^{\text{avg}}(k-1) - 0.4)}{G_{\text{eug-lower}} - G_{\text{hypo}}} + 0.4, & G_{\text{hypo}} \leq G_{\min} \leq G_{\text{eug-lower}} \\ A_f^{\text{avg}}(k-1), & G_{\text{eug-lower}} \leq G_{\min} \leq G_{\text{eug-upper}} \\ \frac{(G_{\min} - G_{\text{eug-upper}}) \cdot (1.3 - A_f^{\text{avg}}(k-1))}{G_{\text{hyper}} - G_{\text{eug-upper}}} + A_f^{\text{avg}}, & G_{\text{eug-upper}} \leq G_{\min} \leq G_{\text{hyper}} \\ 1.3, & G_{\text{hyper}} \leq G_{\min} \end{cases} \quad (5)$$

If $G_{\min}$ is within the target range of $G_{\text{eug-lower}} = 90 \text{ mg dl}^{-1}$ to $G_{\text{eug-upper}} = 140 \text{ mg dl}^{-1}$ (where $G_{\text{eug-lower}}$ and $G_{\text{eug-upper}}$ are the lower and upper limits of a euglycaemic glucose target range, respectively), then the aggressiveness factor does not change and $A_f = A_f^{\text{avg}}$. However, if $G_{\min}$ drops below $G_{\text{eug-lower}}$, the aggressiveness factor, A_f , is reduced proportionally down to a hypoglycaemia threshold of G_{hypo} (70 mg dl^{-1}). Below the hypoglycaemia threshold (G_{hypo}), $A_f = 0.4$, which means that the pre-meal insulin will be dosed at 40% of the original amount as shown in equation (5). In a similar manner, the aggressiveness factor is increased proportionally with respect to $G_{\min}$ if $G_{\min}$ is above the upper limit of the target range ($G_{\text{eug-upper}}$) until $G_{\min}$ exceeds the hyperglycaemic threshold (G_{hyper}). Above G_{hyper} the value of the insulin aggressiveness factor is 1.3. To ensure that the aggressiveness factor accurately reflects user glycaemic response, a minimum of five boluses must be observed within a specific window of time before a new aggressiveness factor is calculated.

Adjustment of carb:ins and correction factors occurs under two separate scenarios. The ALPHA algorithm is used to calculate the precise dosage adjustment to carb:ins only when indicated by the KNN classification procedure. In contrast, the ALPHA algorithm adjusts correction factors directly because the KNN classification does not address correction factors. The maximum adjustment to fast-acting insulin is constrained to $\pm 15\%$ per week. For adjustments to long-acting basal insulin, the dosage is adjusted by $\pm 10\%$ per week when indicated by the KNN classification procedure. The KNN-DSS system accounts for the extended ($>5 \text{ d}$) pharmacologic steady state of ultra-long-acting Tresiba (degludec) and Toujeo (glargine U300) by constraining basal recommendations to one basal insulin recommendation every 2 weeks (Extended Data Figs. 2 and 3). These constraints are a safety measure. Titrations to basal insulin are applied uniformly to all basal doses that may occur at different times of day. For example, for people who require twice-daily basal insulin injections, if a 10% reduction in basal insulin is recommended, both the morning and evening insulin will be reduced by 10%.

Quality control algorithm. A quality control algorithm was developed to ensure that recommendations delivered to the person adhere to physician standards. This algorithm incorporates expert knowledge based on physician input on titration of basal and bolus insulin regimens to ensure that engine recommendations are consistent with physician standards and are safe for a person with T1D. Quality control metrics for each recommendation delivered by the KNN-DSS are shown in Fig. 1d and are elaborated further in Extended Data Figs. 1–7.

Clinical study data and physician review. Data were obtained from 25 people with T1D who participated in a 4-week, outpatient study of CGM-augmented MDI therapy (Supplementary Table 1: Study 1). After data collection techniques were optimized on the first 9 participants, the remaining 16 participants received recommendations for dosing and behavioural changes on the basis of suggestions from the KNN-DSS system. At the end of each study week, an endocrinologist reviewed the data and identified one or more adjustments to insulin therapy. Study participants were equipped with a Dexcom G6 sensor and Apple Watch to track glucose trends and physical activity. Participants used either long-acting Lantus or Tresiba insulin that was captured automatically using the Bluetooth-enabled Gocap or Clipsulin insulin dose-capture devices. Participants used fast-acting Novolog (aspart) insulin captured during the study automatically with an InPen device. Participants were instructed to log meals and exercise using a custom food and exercise tracking app, and to log dose insulin using the InPen app. Out of the 25 participants, 15 were female, the mean weight was $82.73 \pm 19.56 \text{ kg}$, and the mean height was $170.60 \pm 19.56 \text{ cm}$. The mean duration of diabetes was 15.52 ± 6.92 years, the mean age was 30.50 ± 5.92 years, and the mean HbA1c was $8.78\% \pm 1.36\%$. Additional information regarding study population characteristics, recruitment, and ethics oversight can be found in the Reporting Summary.

The study concluded with a total of 78 physician review sessions that accounted for over 500 d of data from 25 participants. For each week of study data, physicians were instructed to identify one or more adjustments to insulin therapy from a set of 12 potential recommendations (Table 1). Data obtained during this clinical study were retrospectively analysed by the KNN-DSS to generate recommendations and calculate physician agreement metrics according to Supplementary Table 10.

Safety review. Recommendations generated by the KNN-DSS for human participants underwent safety evaluation by faculty at the Department of Endocrinology, Harold Schnitzer Diabetes Health Center. Safety of the KNN-DSS recommendations was assessed by having the physician determine whether the recommendation had the potential to cause hypoglycaemia or night-time hypoglycaemia events. A total of 100 safety reviews were performed.

Missing data. At least four days of cumulative CGM data are required by the KNN-DSS framework to provide new recommendations. The KNN-DSS engine framework will refrain from providing a recommendation until sufficient CGM data are present. To address issues of missing data and data misclassification of insulin boluses that are common in real-world datasets, we developed an auxiliary insulin bolus estimation tool. Insulin boluses recorded by the Bluetooth-enabled insulin capture devices were first paired to announced meal entries. Boluses that were not within 20 min of an announced meal were counted as unlabelled boluses. For each unlabelled bolus, we evaluated the glucose level, glucose trend, and correction factor setting at the time of bolus administration. We then estimated the correction insulin dose that would have been called for by inputting this information into the Scheiner trend adjustment calculator³⁴. Any remaining units of insulin are counted as a meal bolus. These estimated contributions are then combined with existing insulin data and are input to the algorithm.

Assessment of the accuracy of KNN-DSS recommendations as compared to endocrinologist recommendations. Glycaemic outcomes and insulin dosing behaviours were analysed by one of three board-certified endocrinologists and by the KNN-DSS. Recommendations delivered by the KNN-DSS were compared to physician recommendations for each week of data collected during the clinical study. Similarity between KNN-DSS recommendations and physician

recommendations were calculated using a modified Sørensen–Dice similarity coefficient²⁵.

$$DSC = \frac{|R_{\text{engine}} \cap R_{\text{physician}}|}{|R_{\text{engine}}|} \quad (6)$$

In equation (6), the similarity between physician and engine recommendations is calculated as the number of recommendations common to both sets, divided by the total number of recommendations delivered by the engine.

Recommendations were classified into one of three categories: ‘agreement’, ‘disagreement’ or ‘additional treatment’ (Supplementary Table 10). Agreement refers to an engine recommendation that was in full agreement with the physician recommendation (a perfect categorical match), or that was in partial agreement with the physician recommendation and titrates insulin in the same direction (for example, different categorical recommendations that both increase insulin). Disagreement refers to an engine recommendation that was in full disagreement with the physician (for example, one recommendation increases basal insulin and the other decreases basal insulin), or that partially disagrees with the physician and titrates overall insulin in a different direction (for example, one recommendation increases meal insulin and the other decreases basal insulin). Additional treatment refers to a scenario in which the engine recommended insulin dosage adjustments, but the physician indicated no change to the settings of the study participant, and vice-versa. Short-acting insulin bolus treatments reflect a 4-h pharmacokinetic activity; therefore, insulin doses in adjacent treatment windows are highly correlated and are considered to be in partial agreement. A behavioural recommendation to be more adherent to a dosing regimen was counted as safe and in agreement. In some scenarios in which the engine recommended to use the bolus calculator and the physician recommended to increase or decrease basal insulin, the recommendations are not comparable. The overall accuracy was obtained by calculating the average of the similarity across all recommendations.

Inter-physician recommendation agreement. Using a dataset³⁰ collected during a one-month outpatient clinical study of open-loop insulin therapy, three board-certified endocrinologists separately reviewed participant CGM data and dosing behaviours and recommended one or more adjustments to insulin therapy (Supplementary Table 1: Study 2). The physicians then collectively reviewed participant data to reach a consensus on what recommendation should be given. The Sørensen–Dice coefficient (equation (6)) was then used to determine the agreement between individual physician recommendations, as well as the accuracy of those recommendations, as compared to the consensus (Supplementary Table 2). A summary of dataset description and usage is available in Supplementary Table 1.

In silico evaluation. We evaluated the KNN-DSS during two in silico studies. In the first study, 29 virtual patients from the OHSU T1D simulator participated in a 75-week study in which the virtual patients used the decision support system weekly to adjust doses of basal insulin, mealtime insulin, and correction insulin. After 52 weeks, we changed certain insulin settings and dosing behaviours and monitored the ability of the engine to recover these settings. In the second study, 100 virtual patients from the UVA–Padova simulator participated in a similar 52-week study of engine usage. To simulate an MDI population using the UVA–Padova simulator, we replaced the default time-varying basal rate, which is characteristic of programmable insulin pumps, with a constant basal dosage that could be titrated weekly by the KNN-DSS.

Virtual patients from both studies exhibited inter-individual variations in weight, total daily insulin requirement, and circadian insulin sensitivities as described above. For both the OHSU T1D and UVA–Padova simulators, patients were fed real-world meal scenarios (Supplementary Table 1: Study 3) that ranged from 2–9 meals per day and occurred at varying intervals. In addition, we imposed errors in insulin dosing settings and adherence to dosing strategies (see Supplementary Table 4), as well as statistical variation in estimation of meal amounts to reflect realistic use of bolus calculators, interstitial glucose CGM measurement noise and circadian variation in insulin sensitivity to reflect realistic glycaemic profiles. We evaluated study outcomes of percent time in target range and percent hypoglycaemia at time points of 1 week, 1 month and 3 months, as well as at the study conclusion.

The virtual patients used for evaluation and a subset of the real-world meal scenarios were excluded from the training process. In this way, performance was analysed on virtual people with T1D that had not been observed before by the KNN-DSS.

Analysis and statistical power. The glycaemic outcomes of percent time in target range and percent time in hypoglycaemia were determined for the virtual patients at each time point of the study. The percent change was calculated across each week of the study compared with the first week of the study, before any recommendations were given. Results are reported by mean and s.d. for normally distributed outcomes, and median and interquartile range for non-parametric data. A students two-tailed, paired *t*-test of $\alpha = 0.05$ was used to determine the significance in the change of glycaemic outcomes, and a two-tailed Wilcoxon

signed-rank test was used to determine significance for non-parametric data. Cohen’s *d* effect size for paired samples was calculated to account for the influence of the in silico framework³⁵ and large sample sizes on *P* value statistics.

Use of human participants. All participants were adults enrolled under informed consent. The pilot study was approved by the Institutional Review Board at OHSU, and additional information can be found at <https://clinicaltrials.gov> under registration number NCT03443713.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data generated in silico during this study and the code used for analysis is available from the corresponding author on reasonable request. Access to human participant data was granted for the current study, and further human data usage or sharing is subject to restrictions and is not publicly available. Requests for restricted, de-identified data on human participants can be submitted to the corresponding authors at OHSU. Requests will be assessed on a case-by-case basis, and are subject to a formal Repository Sharing Agreement. Additional reported outcomes of human participants can be found at <https://clinicaltrials.gov> under registration number NCT03443713.

Code availability

The code used to generate in silico data for this study, the OHSU virtual patient population simulator code, is available at https://github.com/petejacobs/T1D_VPP. Access to the licensed software for the UVA–Padova virtual population was granted for the current study and it can be requested from the developers of this software directly at the University of Padova.

Received: 25 September 2019; Accepted: 23 April 2020;

Published online: 1 June 2020

References

- Paris, C. A. et al. Predictors of insulin regimens and impact on outcomes in youth with type 1 diabetes: the SEARCH for diabetes in youth study. *J. Pediatr.* **155**, 183–189.e1 (2009).
- Miller, K. M. et al. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. *Diabetes Care* **38**, 971–978 (2015).
- Resalat, N., El Youssef, J., Tyler, N., Castle, J. & Jacobs, P. G. A statistical virtual patient population for the glucoregulatory system in type 1 diabetes with integrated exercise model. *PLoS One* **14**, e0217301 (2019).
- Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* **13**, 21–27 (2006).
- Nimri, R. et al. Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: Variations among countries and physicians. *Diabetes Obes. Metab.* **20**, 2458–2466 (2018).
- Man, C. D. et al. The UVA/PADOVA type 1 diabetes simulator: new features. *J. Diabetes Sci. Technol.* **8**, 26–34 (2014).
- The Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group. Intensive diabetes treatment and cardiovascular outcomes in type 1 diabetes: the DCCT/EDIC study 30-year follow-up. *Diabetes Care* **39**, 686–693 (2016).
- Schwartz, F. L., Guo, A., Marling, C. R. & Shubrook, J. H. Analysis of use of an automated bolus calculator reduces fear of hypoglycemia and improves confidence in dosage accuracy in type 1 diabetes mellitus patients treated with multiple daily insulin injections. *J. Diabetes Sci. Technol.* **6**, 150–152 (2012).
- Roze, S. et al. Cost-effectiveness of continuous subcutaneous insulin infusion versus multiple daily injections of insulin in type 1 diabetes: a systematic review. *Diabet. Med.* **32**, 1415–1424 (2015).
- McNally, K., Rohan, J., Pendley, J. S., Delamater, A. & Drotar, D. Executive functioning, treatment adherence, and glycemic control in children with type 1 diabetes. *Diabetes Care* **33**, 1159–1162 (2010).
- Sarbacker, G. B. & Urteaga, E. M. Adherence to insulin therapy. *Diabetes Spectr.* **29**, 166–170 (2016).
- Kirwan, M., Vandelanotte, C., Fenning, A. & Duncan, J. M. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. *J. Med. Internet Res.* **15**, e235 (2013).
- Charpentier, G. et al. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). *Diabetes Care* **34**, 533–539 (2011).
- Wu, Y. et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR Mhealth Uhealth* **5**, e35 (2017).

15. Veazie, S et al. Rapid evidence review of mobile applications for self-management of diabetes. *J. Gen. Intern. Med.* **33**, 1167–1176 (2018).
16. Beck, R. W. et al. Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: the DIAMOND randomized clinical trial. *JAMA* **317**, 371–378 (2017).
17. Steil, G. M. et al. Use of automated clinical decision support (CDS) to effect glycemic control in elderly patients with T1D. *Diabetes* **67**, 921–P (2018).
18. Palerm, C. C., Zisser, H., Jovanovic, L. & Doyle, F. J. 3rd A run-to-run control strategy to adjust basal insulin infusion rates in type 1 diabetes. *J. Process Control* **18**, 258–265 (2008).
19. Herrero, P., Bondia, J., Gimenez, M., Oliver, N. & Georgiou, P. Automatic adaptation of basal insulin using sensor-augmented pump therapy. *J. Diabetes Sci. Technol.* **12**, 282–294 (2018).
20. Toffanin, C., Messori, M., Cobelli, C. & Magni, L. Automatic adaptation of basal therapy for type 1 diabetic patients: a run-to-run approach. *Biomed. Signal Process. Control* **31**, 539–549 (2017).
21. Zisser, H., Palerm, C. C., Bevier, W. C., Doyle, F. J. 3rd & Jovanovic, L. Clinical update on optimal prandial insulin dosing using a refined run-to-run control algorithm. *J. Diabetes Sci. Technol.* **3**, 487–491 (2009).
22. Herrero, P. et al. Advanced insulin bolus advisor based on run-to-run control and case-based reasoning. *IEEE J. Biomed. Health Inform.* **19**, 1087–1096 (2015).
23. Perez-Gandia, C. et al. Decision support in diabetes care: the challenge of supporting patients in their daily living using a mobile glucose predictor. *J. Diabetes Sci. Technol.* **12**, 243–250 (2018).
24. Breton, M. D. et al. Continuous glucose monitoring and insulin informed advisory system with automated titration and dosing of insulin reduces glucose variability in type 1 diabetes mellitus. *Diabetes Technol. Ther.* **20**, 531–540 (2018).
25. Reddy, M. et al. Clinical safety and feasibility of the advanced bolus calculator for type 1 diabetes based on case-based reasoning: a 6-week nonrandomized single-arm pilot study. *Diabetes Technol. Ther.* **18**, 487–493 (2016).
26. Resalat, N., El Youssef, J., Reddy, R., Castle, J. & Jacobs, P. G. Adaptive tuning of basal and bolus insulin to reduce postprandial hypoglycemia in a hybrid artificial pancreas. *J. Process Control* **80**, 247–254 (2019).
27. Sørensen, T. J. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and its Application to Analyses of the Vegetation on Danish Commons (I kommission hos E. Munksgaard, 1948).
28. Davidson, M. B., Duran, P., Davidson, S. J. & Lee, M. Comparison of insulin dose adjustments by primary care physicians and endocrinologists. *Clin. Diabetes* **36**, 39–43 (2018).
29. Bashan, E. & Hodish, I. Frequent insulin dosage adjustments based on glucose readings alone are sufficient for a safe and effective therapy. *J. Diabetes Complicat.* **26**, 230–236 (2012).
30. Reddy, R. et al. The effect of exercise on sleep in adults with type 1 diabetes. *Diabetes Obes. Metab.* **20**, 443–447 (2018).
31. Castle, J. R. et al. Randomized outpatient trial of single- and dual-hormone closed-loop systems that adapt to exercise using wearable sensors. *Diabetes Care* **41**, 1471–1477 (2018).
32. Pettus, J. & Edelman, S. V. Recommendations for using real-time continuous glucose monitoring (rtCGM) data for insulin adjustments in type 1 diabetes. *J. Diabetes Sci. Technol.* **11**, 138–147 (2017).
33. Whitney, A. W. A direct method of nonparametric measurement selection. *IEEE Trans. Comput.* **20**, 1100–1103 (1971).
34. Scheiner, G. *Practical CGM: Improving Patient Outcomes Through Continuous Glucose Monitoring* (American Diabetes Association, 2015).
35. White, J. W., Rassweiler, A., Samhoury, J. F., Stier, A. C. & White, C. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* **123**, 385–388 (2014).

Acknowledgements

The guarantor of this research is Peter G. Jacobs who takes responsibility for the contents of the article. Correspondence and requests for materials can be addressed to Nichole S. Tyler and Peter G. Jacobs. The authors thank Gavin Young for his contributions to algorithm methodology. The authors disclose receipt of the following financial support for the research, authorship, and/or publication of the article: this work was supported by The Leona M. and Harry B. Helmsley Charitable Trust (grant 2018PG-T1D001), the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (grant 1 R01DK120367-01) and a Dexcom grant.

Author contributions

N.S.T., C.M.M.-L., R.H.D. and P.G.J. contributed to the design of the published decision support engine algorithm. N.S.T. and C.M.M.-L. performed the additional evaluations of decision support engine performance. N.S.T., C.M.M.-L., R.H.D., W.W.H. and P.G.J. designed and discussed strategies for engine evaluation and outcomes metrics. L.M.W., J.R.C. and J.E.Y. served as physicians collecting human clinical trial data, contributed to the design of the quality control algorithm, and performed a safety evaluation of the algorithm. D.L.B., V.B.G. and F.H.G. collected and managed the human data used to evaluate the decision support engine.

Competing interests

The authors declare the following competing interests regarding research, authorship and publication of this article: J.R.C. and P.G.J. have financial interest in Pacific Diabetes Technologies Inc. (PDT), a company with potential commercial interests in the results and research of this technology. J.R.C. and P.G.J. are founders and shareholders in PDT and P.G.J. is a board member of PDT. Neither J.R.C. nor P.G.J. receive any financial compensation from PDT as consultants or otherwise, beyond shares in the company. J.R.C. and P.G.J. have received honoraria for consulting and research support from Dexcom. Although the methods on the algorithm were disclosed to the OHSU Technology Transfer Office, there has not yet been a patent filed on the algorithm and PDT does not have any rights to any of the technology described in the paper. N.S.T., C.M.M., R.H.D., L.M.W., D.L.B., V.B.G., F.H.G., W.W.H. and J.E.Y. declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s42255-020-0212-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s42255-020-0212-y>.

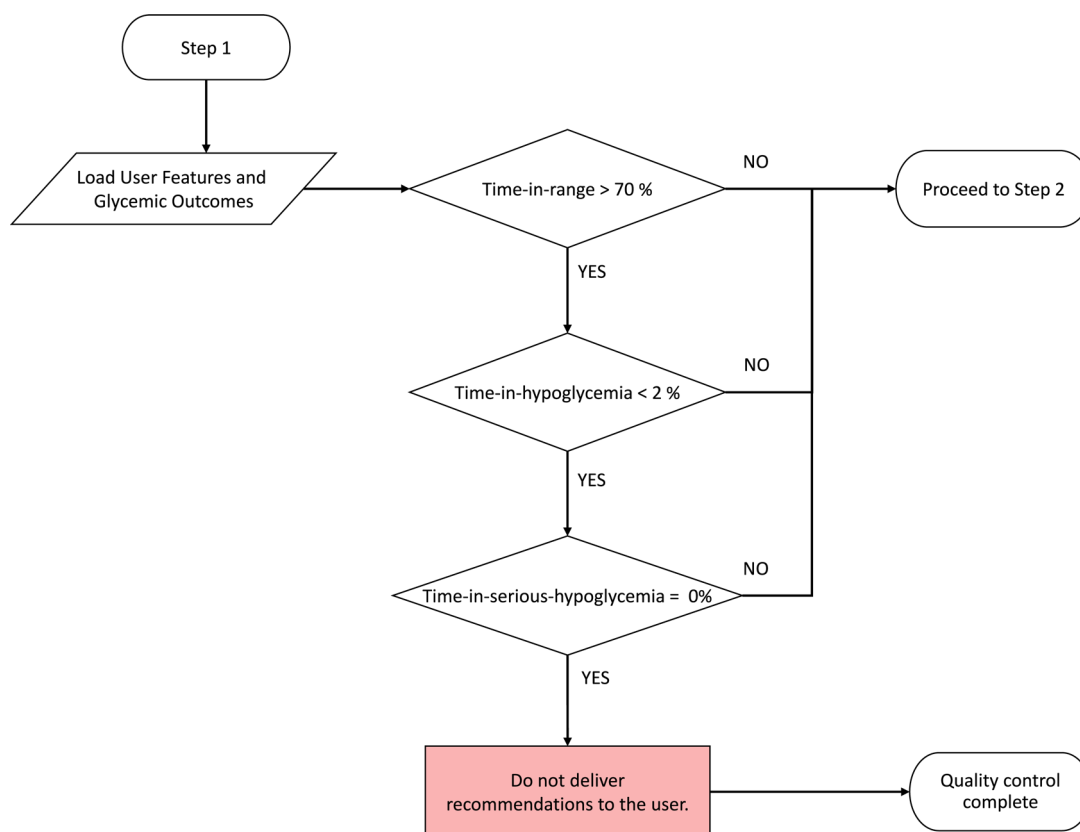
Correspondence and requests for materials should be addressed to N.S.T. or P.G.J.

Peer review information Primary Handling Editor: Christoph Schmitt.

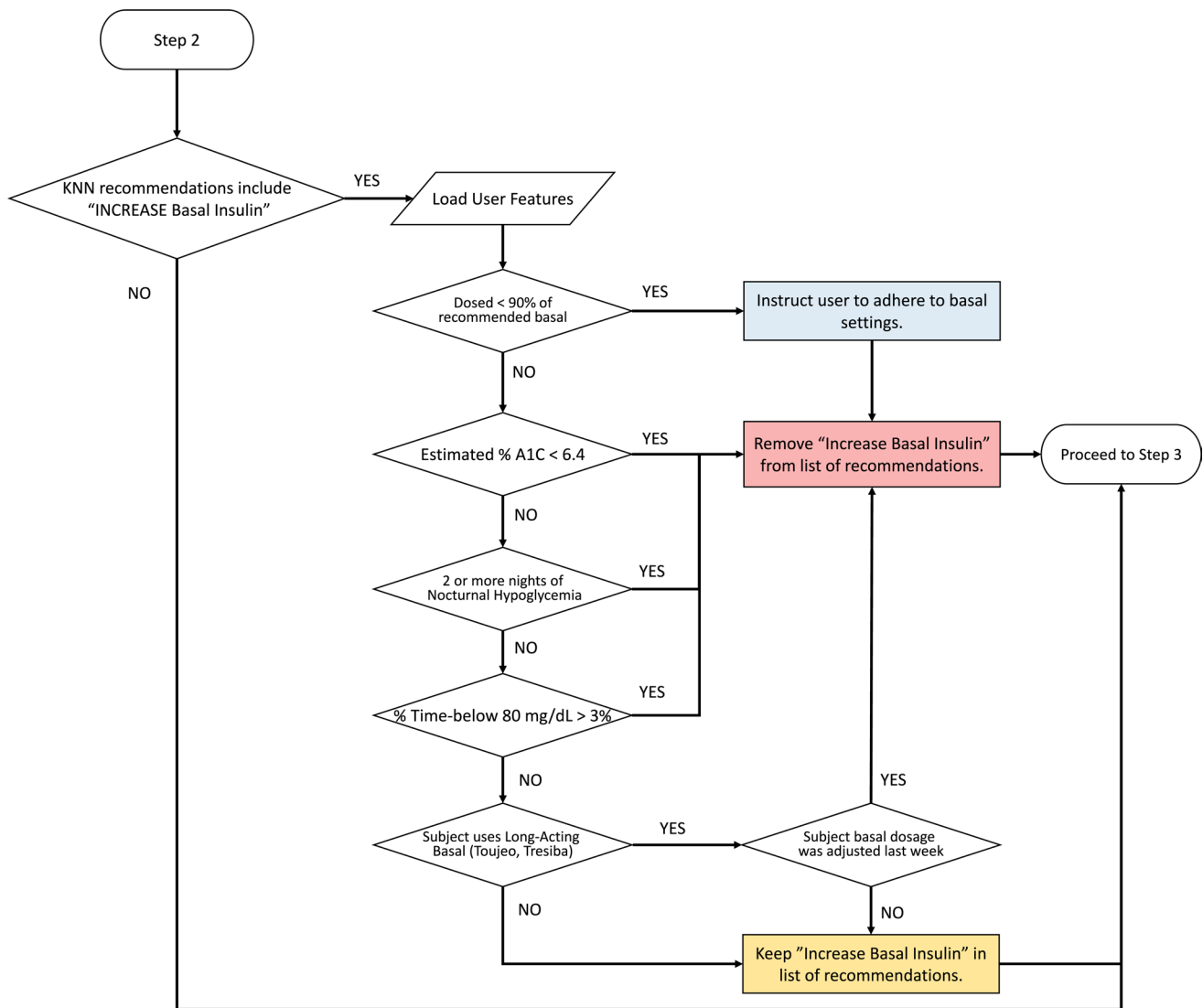
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

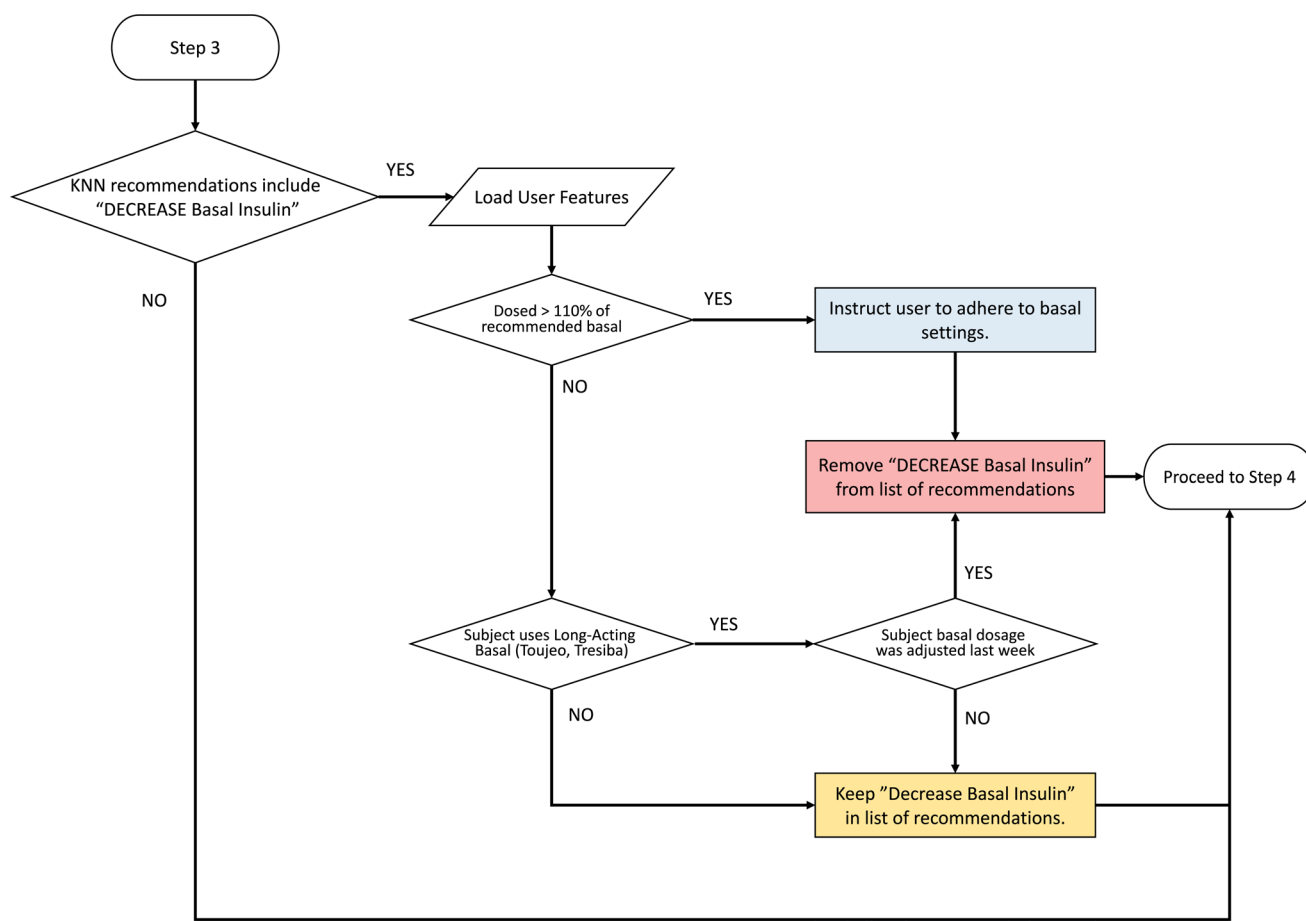
© The Author(s), under exclusive licence to Springer Nature Limited 2020



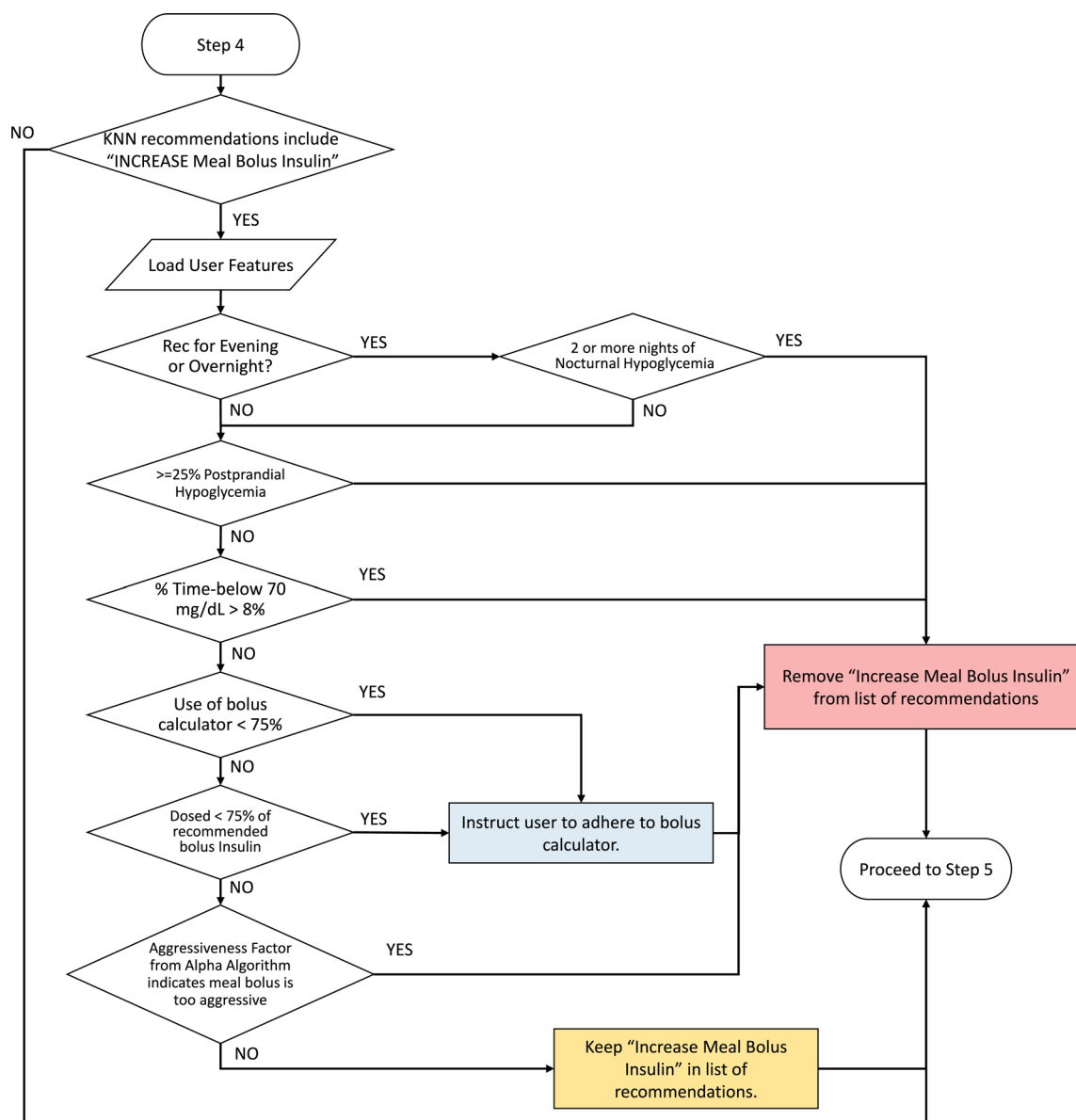
Extended Data Fig. 1 | Quality control algorithm to assess need for insulin titration. Quality control algorithm to assess need for insulin titration. User data and glycemic outcomes are loaded and compared against metrics for percent time in hypoglycaemia, percent time in target range, and percent time in serious hypoglycaemia. If users meet all metrics, recommendations for insulin titration are not required.



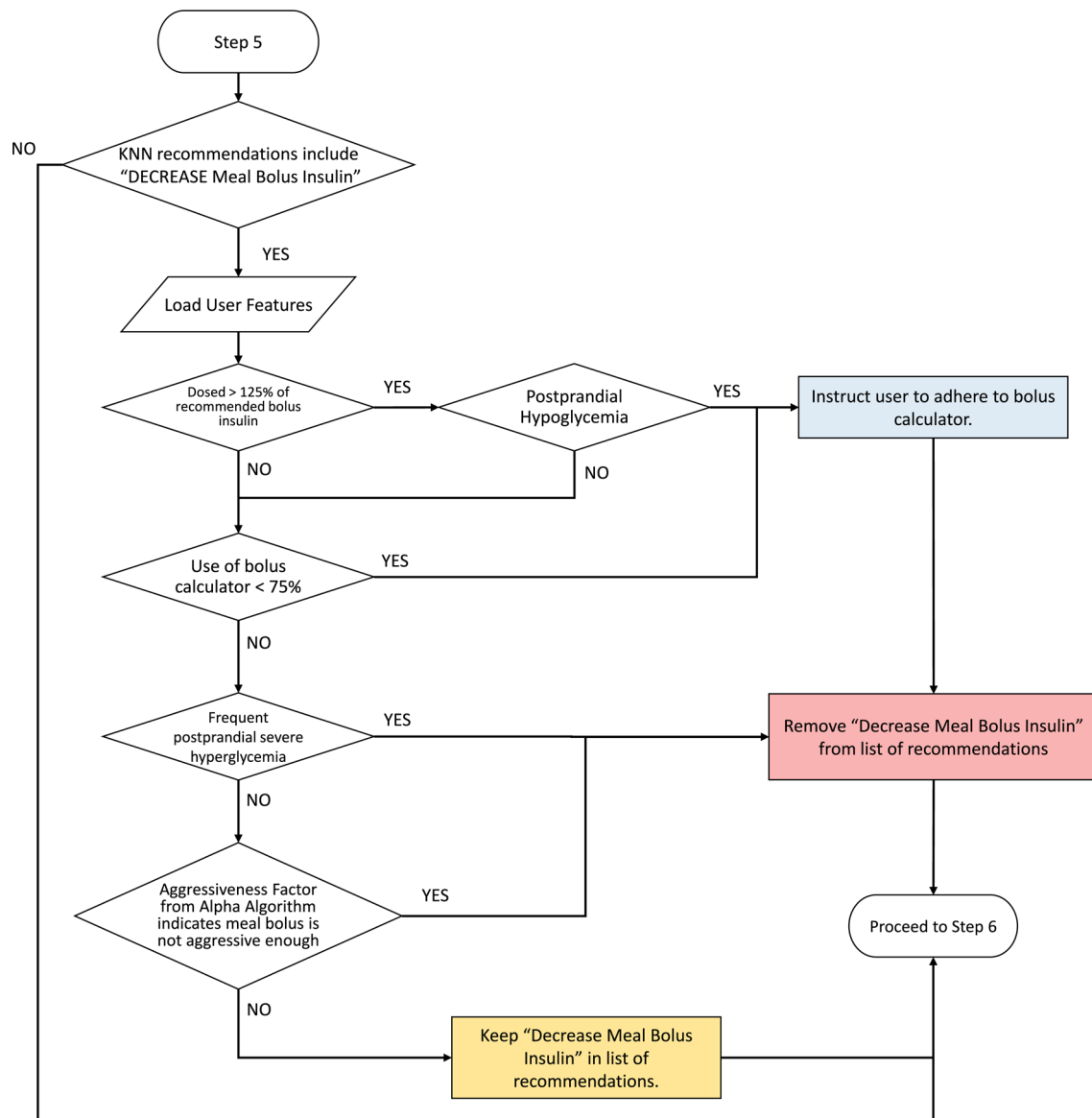
Extended Data Fig. 2 | Quality control algorithm to assess increasing basal insulin dosage. Quality control algorithm to assess increasing basal insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for physician-informed metrics of nocturnal hypoglycaemia, near hypoglycaemia episodes, subject time in target range, subject adherence, and insulin formulation-dependent requirements.



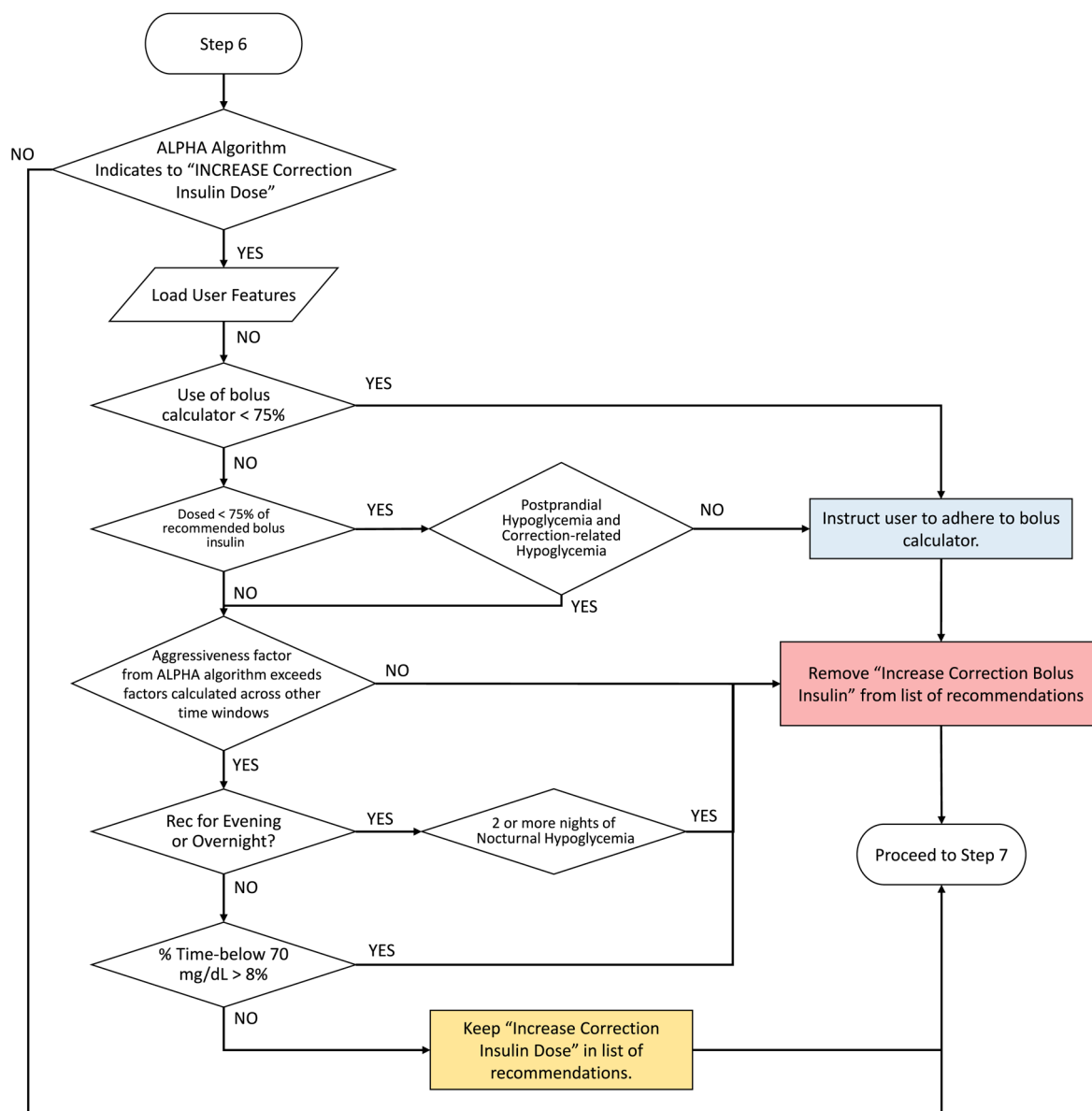
Extended Data Fig. 3 | Quality control algorithm to assess decreasing basal insulin dosage. Quality control algorithm to assess decreasing basal insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for subject adherence, and insulin formulation-dependent requirements.



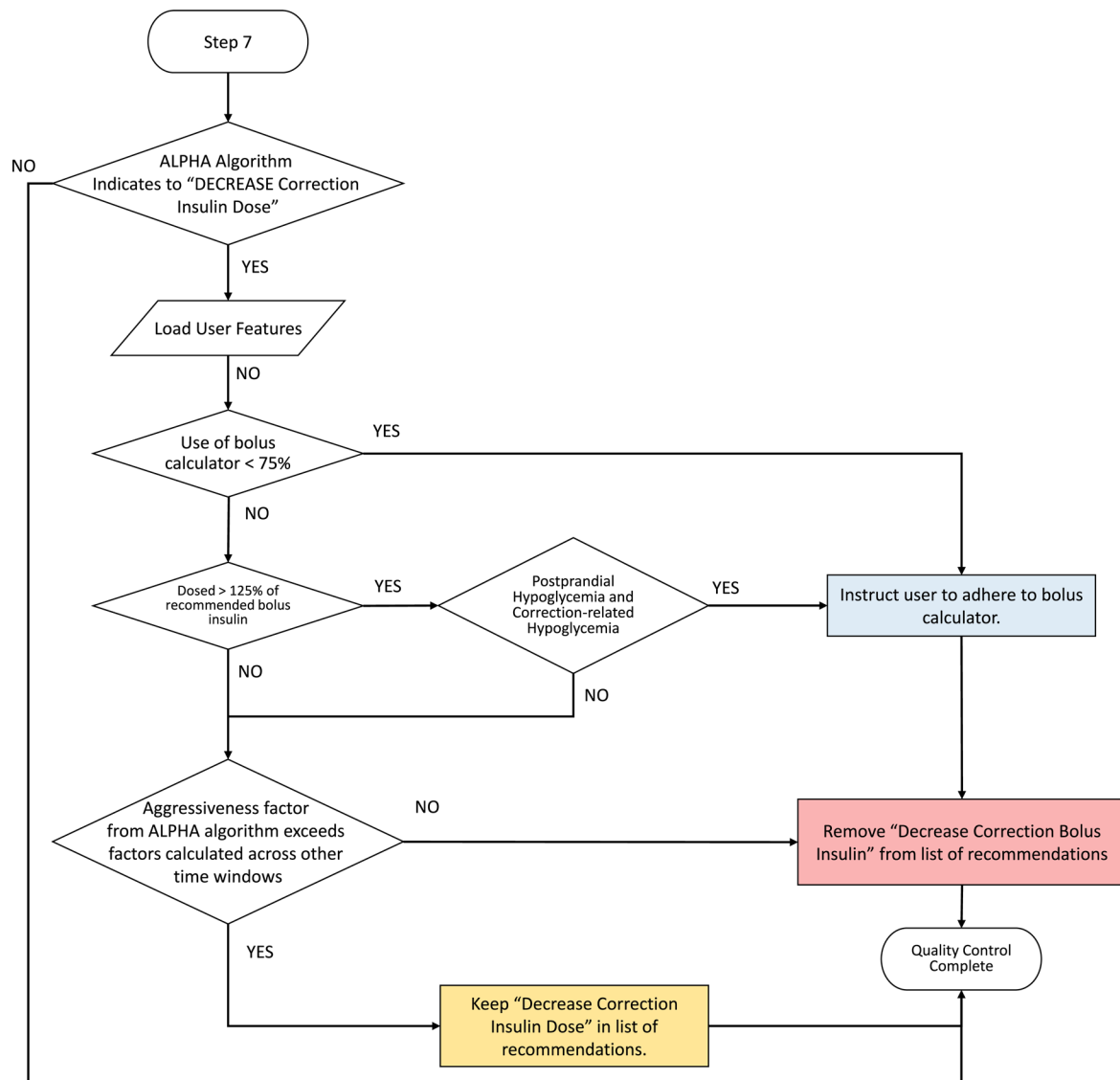
Extended Data Fig. 4 | Quality control algorithm to assess increasing meal bolus insulin dosage. Quality control algorithm to assess increasing meal bolus insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for physician-informed metrics of postprandial hypoglycaemia, subject adherence, and factors returned by the ALPHA algorithm.



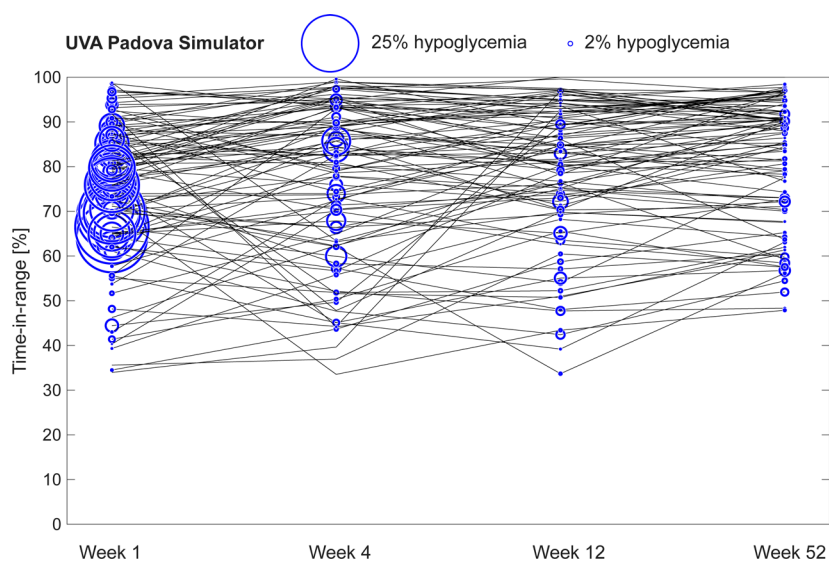
Extended Data Fig. 5 | Quality control algorithm to assess decreasing meal bolus insulin dosage. Quality control algorithm to assess decreasing meal bolus insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for physician-informed metrics of postprandial severe hyperglycaemia, subject adherence, and factors returned by the ALPHA algorithm.



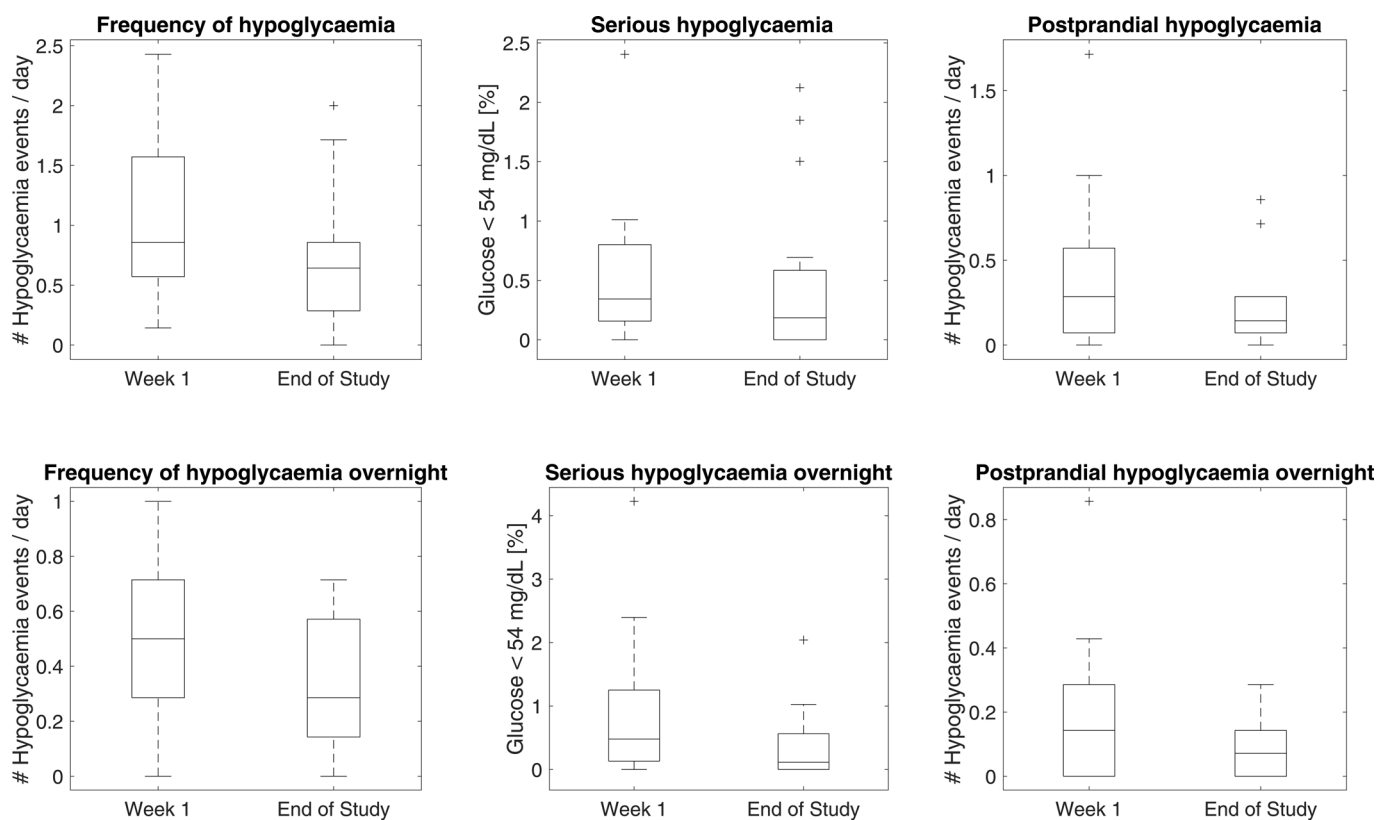
Extended Data Fig. 6 | Quality control algorithm to assess increasing correction bolus insulin dosage. Quality control algorithm to assess increasing correction bolus insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for physician-informed metrics of postprandial and correction-related hypoglycaemia, subject adherence, and factors returned by the ALPHA algorithm.



Extended Data Fig. 7 | Quality control algorithm to assess decreasing correction bolus insulin dosage. Quality control algorithm to assess decreasing correction bolus insulin dosage. User features and glycemic outcomes are loaded by the algorithm and assessed for physician-informed metrics of subject adherence, postprandial and correction-related hypoglycaemia, and factors returned by the ALPHA algorithm.



Extended Data Fig. 8 | KNN-DSS engine performance in improving subject outcomes in an independent virtual patient population. KNN-DSS engine performance in improving subject outcomes in an independent virtual patient population. Glycemic outcomes during a 52-week study of the FDA-approved UVA-Padova virtual patient simulator. Percent time in hypoglycaemia is indicated by the blue circular radius.



Extended Data Fig. 9 | Outcomes of a human pilot study evaluating KNN-DSS augmented decision support. Outcomes of a human pilot study evaluating KNN-DSS augmented decision support over 4 weeks where the first recommendation is given at the start of week 2. For panels a-f, boxplot limits indicate the first and third quartiles, centerline indicates the median, and whiskers mark the last non-outlier data-point within 1.5xIQR. For panels a-f, participant data collected during week 1 and the final week were compared using a two-tailed Wilcoxon signed-rank test, with significance level of 5%. a, Frequency of hypoglycaemia was nominally reduced on the final week compared with week 1 of the study (0.86 vs 0.64, $P=0.051$, $n=16$ independent subjects). b Serious hypoglycaemia was nominally reduced on the final week compared with week 1 of the study (0.34% vs. 0.19%, $P=0.56$, $n=16$ independent subjects). c Postprandial hypoglycaemia events were nominally reduced on the final week compared with week 1 (0.29 vs 0.14, $P=0.08$, $n=16$ independent subjects). d Frequency of overnight hypoglycaemia was significantly reduced on the final week compared to week 1 (0.50 to 0.29, $P=0.04$, $n=16$ independent subjects). e Serious hypoglycaemia overnight was significantly reduced on the final week compared to week 1 (0.48% to 0.11%, $P=0.03$, $n=16$ independent subjects). f Postprandial hypoglycaemia overnight was nominally reduced on the final week compared to week 1 (0.14 to 0.07, $P=0.06$, $n=16$ independent subjects).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data was generated using a virtual patient simulator available in Matlab ver 9.7 and Simulink ver 10.0 softwares. Pilot study data was obtained from human subjects who underwent informed consent (see below, reporting specific materials, systems and methods). Dataset described in Supplementary Information Table 1, Study 1. All other clinical data was obtained previously and was not collected during this study (Supplementary Information Table 1, Study 2 and Study 3).

Data analysis

Data presented in the manuscript was analyzed in Matlab 2019b.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Requests for restricted, de-identified human subjects data can be submitted to Oregon Health & Science University. Requests will be assessed on a case-by-case basis, and are subject to a formal Repository Sharing Agreement. Reported outcomes of all 25 human subjects can be found at [clinicaltrials.gov](#)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Longitudinal data was obtained from 329 virtual patients in two separate virtual simulators. No formal power calculation was performed, and we used the maximum number of subjects available for use. An ad-hoc power calculation using G-Power indicates that a sample size of 30 is required to measure a 10% difference in time-in-range, assuming an expected standard deviation of 18.6% (expected standard deviation from prior decision support studies on MDI subjects). We further analyzed data obtained previously from 25 human study participants during a pilot study. No formal power analysis was performed for pilot study, the primary purpose of the study was data-collection (see clinicaltrials.gov).
Data exclusions	The study reported on clinicaltrials.gov reports 25 subjects. Of those 25, 16 received KNN-DSS decision support. We report the results of 16 human study subjects who underwent weekly KNN-DSS engine decision support. No data was excluded from study subjects who underwent KNN-DSS testing. We did not exclude any subject data in our comparison of KNN-DSS and physicians. Additional information can be found at clinicaltrials.gov (see below).
Replication	N/A. Each subject served as a replicate to evaluate the decision support system described in this paper.
Randomization	N/A. Randomization does not apply to this study. We evaluated the use of the decision support system on all virtual subjects.
Blinding	N/A. Blinding is not relevant to the study. We evaluated the use of weekly decision support and all subjects underwent the same therapy.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	To compare the KNN-DSS recommendations to endocrinologists, we utilized data from 25 human study participants during a pilot study. We additionally evaluated the safety of the KNN-DSS on 16 of these subjects. Out of the 25 participants, 15 were female, the mean weight was 82.73 +/- 19.56 kg, and mean height 170.60 +/- 19.56 cm. The mean duration of diabetes was 15.52 +/- 6.92 years, mean age was 30.50 +/- 5.92 years, and mean A1c was 8.78 +/- 1.36.
Recruitment	Subjects were recruited from the diabetes research center at Oregon Health & Science University from flyers posted in approved places at OHSU or posted on the web to the clinical trials page for the OHSU Schnitzer Diabetes Clinic, to the Clinic's facebook group, electronic newsletter or from the OHSU Subject Recruitment website. Records from OHSU Schnitzer Diabetes Clinic patients were also screened to find potential subjects. Subjects were also recruited from a list of subjects who participated in past OHSU studies who have agreed to be contacted regarding future studies involving from the study PI's, OHSU diabetes research registry and/or www.clinicaltrials.gov .
Ethics oversight	This study was approved by the Institutional Review Board at Oregon Health & Science University, and additional information can be found at clinicaltrials.gov under registration number NCT03443713.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)
All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	NCT03443713
Study protocol	Protocol available at clinicaltrials.gov
Data collection	The pilot study was performed between July 2018 to May 2019. Data was collected on-site at Oregon Health & Science university. Additional information can be found on clinicaltrials.gov
Outcomes	Primary outcomes were percent time-in-range (70-180 mg/dL) and percent time-in-hypoglycemia (<70 mg/dL), as determined on time frame: days 1-7 vs. days 22-28 of the study. Additional information can be found on clinicaltrials.gov